USDA

United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research Division

RD Research Report
Number RDD-00-05

September 2000

Evaluation of the Use of Spatial Modeling to Improve County Yield Estimation

Chadd Crouse

EVALUATION OF THE USE OF SPATIAL MODELING TO IMPROVE COUNTY YIELD ESTIMATION, by Chadd Crouse, Sampling and Estimation Research Section, Research and Development Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250, September 2000, NASS Research Report, Report No. RDD-00-05.

ABSTRACT

Crop county estimates, produced by the USDA's National Agricultural Statistics Service (NASS), are used by the private sector; colleges and universities; and local, state and national governments to monitor shifts in agricultural production. These estimates are based on a non-probability sample of farming operations in a state, so, as Stasny, Goel, et. al. (1995) point out, they cannot be generated using traditional estimation methods based on known selection probabilities.

Statisticians at The Ohio State University, under a cooperative agreement with NASS, have developed a county yield estimation algorithm based on the assumption that neighboring counties have similar yields. The algorithm is a mixed-effects model which uses post-stratification on total land operated and allows for differences in yield by county and farm size.

This report presents the results of a simulation study designed to assess the effectiveness and reliability of the algorithm. The results indicate that the algorithm produces improved yield estimates when compared to standard ratio estimates. This paper briefly discusses the spatial methodology used in the model, explores potential problems with the algorithm, examines the results of preliminary testing and suggests several steps that should be taken in preparation for operational implementation.

KEY WORDS

Crop County Estimates, Non-Probability Sample, Small-Area Estimation, Yield Simulation

ACKNOWLEDGMENTS

The author would like to thank the Michigan field office in general and Sam Bruner in particular for supplying the necessary data files, and for providing valuable feedback during testing; and Wendy Scholetzky and Dale Atkinson for support and guidance during the creation of this report.

TABLE OF CONTENTS

SUMMARY

The National Agricultural Statistics Service (NASS) has produced county crop yield estimates since at least 1917. These estimates are arguably the most requested and some of the most scrutinized ones produced, but the methodology used to produce them has remained essentially unchanged in recent years (Iwig, 1993). The estimates have been based on non-probability samples, so traditional small area estimation techniques based on known selection probabilities cannot be employed (Stasny, Goel, et. al., 1995). Currently, a standard ratio estimator is the basis of most yield estimates, but Stasny, Goel, et. al. at The Ohio State University (OSU) suggest the use of a mixed-effects model which exploits the correlation between neighboring counties.

The staff at OSU have implemented the proposed model in the form of a SAS macro. The macro uses an Expectation-Maximization (EM) algorithm to produce yields based on previous year estimates, the current year reported data for the county of interest and the current year reported data from counties considered neighbors of the county of interest. The model also employs post-stratification based on the reported size of each operation as determined by the total land operated. This model-based estimation approach was developed in the hopes of providing the following benefits:

1) To allow repeatability of the estimation process.
2) To provide some measure of the variation of the estimates.
3) To produce reasonable yield estimates when no survey data are present.

Several potential problems became apparent during initial testing. The most prominent of these was a lack of convergence of the algorithm and concerns about the post-stratification process. To evaluate the impact of these issues, a data set was generated based on the 1997 Michigan Census of Agriculture and 1998 Michigan county estimates.

Rare commodities present special problems in estimating county yields. These problems, which arise from the small number of usable reports, include estimates with extremely large variances and the unavailability of standard ratio estimates. Of the crops for which Michigan produces county estimates, barley has the fewest number of acres harvested, so it was chosen as an initial test commodity to serve as a proxy for rare commodities. The barley data set was used to test the effectiveness of the algorithm, and to determine if the model-based approach achieves the desired goals stated at the outset for a minor commodity. The testing done here indicates that the algorithm does a reasonable job of estimating yield for barley. Testing also indicates that further improvements could be made, particularly concerning the lack of convergence.

A corn data set was generated in the same manner as the barley data set. Corn was chosen because it is the most common crop in the state in terms of acres harvested, and was used to test

the applicability of the algorithm to more common, wide-spread commodities. Again, the model-based approach produced reasonable yield estimates, and the problems present for barley, the lack of convergence in particular, were not apparent for corn. The lack of problems for corn was most likely the result of a larger sample size; however, simply increasing the sample size may not guarantee convergence. In both cases (barley and corn), the model-based estimates were better than the ratio estimates at minimizing the deviation from the true yield, which is defined as the "actual" county yield calculated from the generated population.

Several recommendations are presented which may need to be addressed before wide scale implementation. In particular, it is recommended that:

1) The algorithm should be tested and evaluated in additional field offices on a wide variety of commodities.
2) The necessary procedures to merge the algorithm with the current county estimates summary system should be examined and developed.
3) The future OSU paper concerning the development of the algorithm should be obtained and evaluated.
4) Work by Dr. Dan Griffith at Syracuse University to improve the current algorithm or possibly develop a new estimation method should be evaluated.

Introduction

Crop county estimates are a significant and important data series produced by the USDA's National Agricultural Statistics Service (NASS). They are used by the private sector; colleges and universities; and local, state and national governments to monitor shifts in agricultural production. Iwig (1993) provides a comprehensive description of the estimation process and points out that the process has remained the same for quite some time. A statistician often has somewhat reliable administrative data available when developing county acreage estimates, and the acres planted in a county for a given commodity rarely change dramatically from one year to the next, so deriving *acreage estimates* is a reasonably straight forward and replicable process. County *yield* estimation, however, is often the most difficult step in the county estimates process. With a few notable exceptions, administrative data are rarely available, and yields can vary wildly from one year to the next. In many cases, very little survey data are available to provide a reliable yield indication. Additionally, millions of dollars in crop insurance payments in any given county are based at least partially on NASS yield estimates, so these estimates are extremely important to our data users, and thus are scrutinized more heavily than many other estimates.

The current sampling strategy involves selecting a large stratified sample of farms which have not been selected for any other survey during the previous year. Questionnaires are then mailed to the sampled operations, and a telephone follow-up is conducted which targets counties with specific commodities or with few mail returns. No attempt at accounting for all nonresponse is made. Additionally, field offices collect crop data quarterly using a nationally allocated probability sample. The September and December quarterly samples are merged with the non-probability sample to create a combined data set which is then summarized to produce various estimates for each commodity. In most cases, state-level estimates based on the quarterly probability surveys have already been determined and published. Additionally, each state or region is divided into several Agricultural Statistics Districts (districts) which presumably represent similar agricultural counties. Estimates for these districts are summarized, reviewed and adjusted by subject matter experts. The same procedure is then performed for each county within a district. So the county estimates are constrained to these district and state estimates. Because these county-level estimates are based on this non-probability sample of farming operations, as Stasny, Goel, et. al. (1995) point out, traditional small area estimation methods based on known selection probabilities cannot be used to produce crop yield estimates.

Stasny, Goel and others at OSU, under a cooperative agreement with NASS, developed an improved county yield estimation algorithm based largely on the assumption that neighboring counties have similar yields. The algorithm is a mixed-effects model with size group, determined by total land operated, as a fixed effect and

county of operation as a random effect.

The primary goal of utilizing such an algorithm is to allow for repeatability of the estimation process. Repeatability implies that two statisticians attempting to produce yield estimates for a given county with the same available information will produce essentially identical estimates. As mentioned previously, a significant amount of scrutiny is placed upon these estimates because crop insurance payments are based on the county yield estimates. Because of this scrutiny, the yield estimates must be defensible and repeatability is a key element towards achieving this goal.

Another goal is to provide some measure of the variance of the estimates. The estimate of the variance provides a measure of quality of the indication. The variance is used to define an interval around the estimated yield that has a known probability of containing the "true" yield.

A third goal is to provide reasonable yield estimates for counties with little or no survey data available. Legitimate questions have been raised concerning whether or not estimates should ever be published in these cases, however, this is a decision left to the field office and will not be covered here. The goal of improving estimates for counties with little or no survey data is achieved by using information from neighboring counties. This goal goes hand-in-hand with the goal of repeatability. Often reports include planted and harvested acres with no yields available. In these cases, harvested acre estimates are easily determined but yields are difficult to derive. The lack of yield data can result in different statisticians setting drastically different yield estimates. The yield indications produced by this algorithm will reduce or eliminate the subjectivity in the yield estimation process.

An additional benefit of the program is to reduce the probability of drastic yield differences between neighboring counties. In the past, statisticians often plotted yields on a map to ensure that this goal was achieved. By directly using information from neighboring counties in the models, the program would minimize these large differences unless the currently reported data for the counties clearly support them.

This paper will briefly discuss the spatial methodology used in the model. Additionally, it will cover problems that may arise during use in field offices, examine the results for corn and barley in Michigan, and make recommendations for future improvements including plans for further testing in preparation for operational implementation.

The Algorithm

The algorithm is programmed as a SAS macro and is designed to provide statisticians with an improved yield estimation tool over currently available indications. It is designed to produce more consistent yields than the current ratio estimate regardless of the sample. It is based on the premise that crop yields in one county are similar to the yields for the same commodity in neighboring counties. The

algorithm uses the reports from the county of interest and also assigns some weight to reports from neighboring counties to derive the yield for the county of interest. For this paper, neighboring counties are defined as any counties which share a border of more than a single point. For example, in Figure 1 Oklahoma and New Mexico are considered neighbors, but Colorado and Arizona are not considered neighbors. This definition of neighbors was chosen primarily to eliminate any arbitrary decision making on the part of the author. However, the field office may choose to use a different definition than described above.

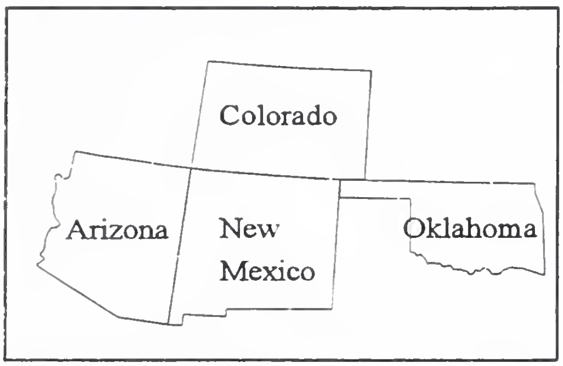

Figure 1 - Neighbors

The current sample data are post-stratified by county and farm size based on total land operated, and yield estimates are derived for each size group in every county. The percentage of census farm acres by size group in a particular county are then used as weights to derive the county yield. The farm size groups that are used can be determined by each field office, but a recommendation will be made here.

The algorithm also uses previous year county yield estimates as starting points for the process. In the event that no previous year estimate was available for a county, the district yield was used, and if no district estimate was available, the state yield was used.

Methodology

In this paper, the effectiveness and reliability of the algorithm is assessed using a simulation approach. The Michigan 1997 Census of Agriculture data and 1998 county yield estimates for the 1997 corn and barley crops were used to generate yields at the record level. These were then used to create a known population large enough for sampling on which the simulations were performed.

Prior to testing the algorithm, some idea of the "true" yield for each county was required. This was determined by merging the record level barley harvested acres reports from the 1997 Michigan Census of Agriculture with the statistician's county level yield estimates for the same crop year. A yield for each of the 619 records with barley was then derived from the distribution generated by Equation 1.

This function was derived by performing regression analysis on 1998 Michigan barley county estimates raw data. The error term (E_{ij}) was a normally distributed random variable generated individually for each record, with the mean and variance based on the distribution of the actual barley data. The size groups are based on total land operated for each operation, and used the

Equation 1. Equation to Derive Barley Yields

$$Y_{ij} = -1.1787 + 0.84407 * Yield_i + 0.45537 * ISize1_j + 2.09019 * ISize2_j + 7.32359 * ISize3_j + E_{ij}$$

where,

Y_{ij} = Derived Yield for Record j in County i.
$Yield_i$ = The 1998 Statistician County Yield from the Michigan County Estimates Files
$ISize1_j$ = 1 if Record j is in Size Group 1 (<180 Total Acres), else 0
$ISize2_j$ = 1 if Record j is in Size Group 2 (180 to 500 Total Acres), else 0
$ISize3_j$ = 1 if Record j is in Size Group 3 (>500 Total Acres), else 0
$E_{ij} \sim N(0, 13)$

original size groups defined by OSU. The equation was also altered somewhat since negative yields for some records occasionally occurred. In these cases, new yields were recalculated until a positive was returned. In order to adjust for bias, 10,000 "populations" were generated with the above equation, and the regression intercept was determined based on the average of these populations. The final population for use in simulation was then created with Equation 1. The state-level yield produced by weighting the modeled yields by harvested acres was 49.96 for the simulated population compared to the official state yield of 50.

A few concerns about the post-stratification process have been raised. First, the percentages of the census farm acres in each size group are used as weights to derive county yields. The Census of Agriculture is conducted every five years with years ending in two or seven serving as the reference year. Using the census data to drive the algorithm is not a problem shortly after census years, but a natural concern is that the quality of these data erode as we move away from the census year. To test the sensitivity of the estimation process to the weight file used, a sample of 200 barley records was selected, and the program was run using the data from both the 1992 and 1997 Census of Agriculture. The average absolute difference for the county yields using the two different base files was only 0.14 bushels per acre. The published estimates are rounded to the whole bushel per acre, so this 0.14 difference is insignificant. Therefore, using 1997 files for post-stratification until the 2002 files are available appears to be a reasonable course of action. However, large changes in county acreage over the five year period may result in significant differences, so additional testing should be completed.

The second concern is the definition of the post-stratification size groups, which are

based on the total acres in an operation. OSU originally used three groups: acres < 180, 180 <= acres < 500 and acres >= 500. These strata were based primarily on the published estimates available from the 1992 Census of Agriculture, and they seem to be logical. However, in the event that a sample contains no records in a given stratum, the macro fails. As a result of the sample design employed, all large operations are included in the sample, so a given sample will probably contain at least one operation in the large stratum. Additionally, in most states the majority of farms (70 percent of Michigan farms) fall in the lower stratum, so samples will almost certainly contain at least one observation in the small stratum. However, there is a reasonable chance that the middle stratum will contain no observations, so different strata definitions may be beneficial in an operational environment.

The population used in the simulation evaluation was derived using the three size groups defined above. For the purpose of evaluating alternative size groupings, the estimates created using the three groups in the algorithm will be considered "correct". Two new group definitions, each with two strata, were created and estimates were compared to those produced using the original strata definitions. The strata definitions are described in Table 1.

In order to determine if one of the new definitions is better than the other (and as effective as the original three stratum definition), 100 samples of 80 operations from the barley population were created. These included all large farms, as the sample would in an operational setting. Yields were generated for each sample using each of the three post-stratification methods. The sum of the squared differences between the modeled county yields from the two new definitions and the original were calculated and compared. In 47 percent of the samples, the sum of squared differences was smaller for New Definition #1 than for New Definition #2, which suggests that neither definition is "better" than the other. However, the average squared difference was 70 percent larger for New Definition #2 than for New Definition #1. The average absolute difference for New Definition #1 was less than 0.7 bushels per acre. Because the average difference is relatively small, and the change to a new definition of strata will result in a reduced probability of non-convergence, converting to New Definition #1 seems to be a reasonable decision. Although the original barley population was generated using the original three stratum

Table 1. Strata Definitions

Post-Stratification Definition	Stratum #1	Stratum #2	Stratum #3
Original	Land < 180	180 <= Land < 500	500 <= Land
New #1	Land < 180	180 <= Land	
New #2	Land < 500	500 <= Land	

definition, all subsequent macro executions described in this paper were run with this new two stratum definition. This approach should help to determine whether or not an incorrect choice of size groups in the algorithm parameters dramatically influences the results.

Potential Problem

The primary problem that will likely arise while using the SAS macro in field offices is an occasional lack of convergence to a solution for all counties. The algorithm is designed to run until the relative group (gpdist) and log-likelihood (lldist) distances reach a preset limit. After each iteration of the algorithm, the yields are adjusted slightly with the two functions normally approaching zero with each iteration. The algorithm stops when the functions reach the tolerance levels or the number of iterations reach a pre-determined threshold. However, when there are very few available reports for a commodity, convergence may not occur after a reasonable number of iterations. If the algorithm does not converge, yield estimates are usually produced, but they may not be "good" ones. This may cause commodity statisticians to discount not only the "bad" yields, but ALL yields produced with the algorithm.

There were only 80 barley reports in the 1998 Michigan county estimates data file, which is the smallest number of any commodity for which estimates would be set. So, to test the potential for a lack of convergence, Michigan barley estimates were examined. For this test, 100 random samples of 80 operations were processed through the system, and the number of iterations until convergence was counted. The sample size of 80 was chosen to match the number of barley reports in the 1998 county estimates data. A maximum of 5000 iterations was allowed for this test. Table 2 lists the results.

The table shows that even with a very small sample the macro usually converges within a reasonable number of iterations. The row labeled "> 5000, Ests" are those cases in which convergence was not reached within 5000 iterations, but yield estimates were produced. The row labeled "> 5000, No Ests" are those cases for which convergence

Table 2. Iterations Required for Convergence of the Algorithm

Iterations	Number of occurrences	Cumulative total
< 100	17	17
100 to 500	53	70
500 to 1500	8	78
1500 to 5000	1	79
> 5000, Ests	12	91
> 5000, No Ests	9	100

was not reached and no yield estimates were produced.

The cases in which convergence was not reached and no yield estimates were produced are of particular interest. A sample for which convergence was not achieved in 5000 iterations was chosen and examined in more detail. In fact, the algorithm for this particular sample did not converge in 50,000 iterations. However, the removal of two specific observations resulted in convergence after only 37 iterations. These two observations were significantly larger than the other yields in the county, and the county also was the second largest county in the sample in terms of acres harvested. It is generally not desirable to delete observations from a sample, but the alternative, no estimates at all, is clearly not desirable. So, prior to wide adoption of the algorithm a procedure should be developed which identifies and removes "problem" records from the sample when convergence is not reached.

Examination of Results for Barley

Since the purpose of this macro is to produce better county yield estimates, no evaluation of it is complete without an examination of the resulting county yields. In order to test how well the model achieves the goals outlined in the Introduction, the derived barley population discussed in the Methodology section was used to generate estimates for the 100 samples of 80 operations. The generated estimates included OSU model estimates and ratio estimates. The ratio estimates were simply the sum of reported production divided by the sum of reported harvested acres. Operationally, these ratio estimates are generated using a form of stratified sampling, and are thus weighted by the sampling rate. However, the weighting is not easily replicated, and there was no significant difference between the weighted ratio and non-weighted ratio yield estimates for the 1997 Michigan barley county yields (p-value = 0.42). Therefore non-weighted ratio yield estimates are used here.

Note that the operational county estimates sample will include observations for all large farming operations. Also, the office staff is able to target certain counties or commodities during data collection to ensure adequate coverage of all areas of a particular state or region. The samples chosen here are simple random samples of the entire population, so this test can be considered the worst case scenario. Therefore, the results should be better in practice.

Any counties with fewer than three operations in the population were excluded from the comparisons below, which resulted in 26 of the 83 counties dropping from the study. These 26 counties were excluded because NASS data disclosure rules prohibit publication when fewer than three observations are present. Also, the RMSE for the ratio estimates was not representative of the true error when these counties were included. For example, if a county contained only one record, the RMSE for that county would equal the yield if the record was not sampled and zero if the record was sampled. To place both

Table 3. Comparison of Barley Estimates

Parameter	Model Estimates	Ratio Estimates
R^2	0.799	0.920
Weighted R^2	0.882	0.957
Sum(RMSE)	420	533
Estimates Produced (%)	62.4	67.6
Better Outliers (%)	49.1	3.5
Sum(Model RMSE Ests)	444	--------

estimation techniques on even ground, these counties were excluded from further analysis.

Table 3 displays some of the key results of the simulation study. The "Model Estimates" are the estimates derived from the OSU algorithm. The two R^2 estimates are the average of the 100 sample estimates of R^2 in which the sample yields were compared to the "actual" population yields. The "Weighted R^2" are weighted by the harvested acres for each county. The estimates produced for Sum(RMSE) and the two R^2 values are based only on the counties in which both a ratio yield and an OSU yield are produced. There were 2429 cases which met this criterion. It should be noted that this criterion penalizes the OSU program by ignoring a key feature of the algorithm which is the production of estimates when no data are present. The "Estimates Produced (%)" is the percentage of times an estimate was produced and is based on the 5700 (100 samples times 57 counties) possible estimates. The "Better Outliers (%)" is the percent of the 57 counties in which the minimum and maximum estimates are both closer to the "True Yield" for the estimator of interest. The "Sum(Model RMSE Ests)" is the sum, across the 57 counties, of the average model produced estimates. This statistic, produced independent of the simulation approach used here, will be discussed in more detail later in this section.

The R^2 values and "Estimates Produced (%)" indicate that the ratio estimates are "better" than those produced by the model. However, the relatively low values of these statistics for the model estimates are simply a result of the OSU program's failure to converge to a solution for a significant number samples. As discussed previously, if a method is developed which will minimize or eliminate this problem, the program may produce better results. The Sum(RMSE), which is a key element, does indicate that the model estimates may be better than the ratio estimates. Another positive attribute for the model is that the outliers are not as extreme.

To determine if one estimator is significantly closer to the true yield than the other, the squared errors for both the model estimates

and the ratio estimates were calculated for each sample in every county. The squared error for the ratio estimates was then subtracted from the squared error for the model estimates, and t-tests were performed on these differences for each of the 57 counties in the study. It was hypothesized that the model yields are better than the ratio yields, so the test took the following form:

H_o: Difference = 0

vs.

H_a: Difference < 0

In this case, 34 of the 57 county differences (60%) were significantly less than zero at the $\alpha = 0.05$ level. Also, only 6 (11%) county differences were significantly larger than zero.

Overall, it cannot be stated with certainty that the current OSU program is significantly better or worse than the ratio estimates for barley in Michigan.

Finally, the model produces an estimate of the RMSE for each model yield. It is interesting to compare these model-produced RMSE estimates to the sample RMSE estimates calculated for this research evaluation. The Sum(Model RMSE Ests) found in Table 3 is reasonably close to the Sum(RMSE) which was based on the repeated sample simulation approach. The correlation between the average sample RMSE and the average model RMSE is 0.365 which indicates a positive correlation, as desired. These statistics indicate that the model does a reasonable job of error estimation. These error statistics are based on samples for which convergence was reached in counties with three or more reports in the known population.

Table 5 in the Appendix features a complete listing of various statistics for every county. Table 6 features complete statistics for the hypothesis tests.

Examination of Results for Corn

In order to determine if the macro performs well for prevalent, wide-spread commodities, corn in Michigan was examined. A corn population was derived in much the same manner as the barley population. Again, this was done by merging the record level corn harvested acres reports from the 1997 Census (16,712 reports) with the statistician's county yield estimates for the same crop year. A yield for each record was then derived from the distribution described in Equation 2.
The distribution was determined by performing regression analysis on record-level data from the 1998 Michigan county estimates survey. The error term (E_{ij}) was a normally distributed random variable generated individually for each record, with the mean and variance based on the distribution of the actual corn data. The size groups are the two size groups discussed in the Methodology section. The equation was again altered somewhat because the large variance on the error term occasionally resulted in negative yields for counties. In these cases, new yields were recalculated until a positive one was returned. To adjust

Equation 2. Equation to Derive Corn Yields

$$Y_{ij} = 5.4934 + 0.940143 * Yield_i - 10.495857 * ISize1_j + 2.519122 * ISize2_j + E_{ij}$$

Where,

Y_{ij} = Derived Yield for Record j in County i.
$Yield_i$ = The 1998 Stat Yield from the Michigan County Estimates Files
$ISize1_j$ = 1 if Record j is in Size Group 1 (<180 Total Acres), else 0
$ISize2_j$ = 1 if Record j is in Size Group 2 (Total Acres >= 180), else 0
$E_{ij} \sim N(0, 26)$

for bias, 10,000 "populations" were generated with Equation 2, and the regression intercept was determined based on the average of these populations. The final population was then created with Equation 2. The state-level yield produced by weighting the modeled yields by harvested acres was 110.9 for the simulated population compared to the official state yield of 111 bushels per acre.

Once again, in order to determine how well the macro achieves the goal of producing consistent estimates, 100 samples were selected from the derived population. Each sample consisted of 1500 records, which represents the approximate number of records with corn in the 1998 Michigan county estimates survey. Additionally, any county with fewer than three operations in the population were excluded from the comparisons below, which resulted in 10 of the 83 counties dropping from the study. These ten counties were excluded because NASS data disclosure rules prohibit publication when fewer than three observations are present.

Table 4 displays some of the key results for corn. The "Estimates Produced (%)" again is the percentage of times an estimate was produced and is based on the 7300 (100 samples times 73 counties) possible estimates. The "Better Outliers (%)" is the percent of the 73 counties in which the minimum and maximum estimates are both closer to the "True Yield" for the estimator of interest. And, once again, only cases for which both a ratio estimate and a model estimate were produced are considered in the correlation and RMSE comparisons. There were 6767 cases which satisfied this criterion.

Some observations from these data are apparent. First, the ratio estimates were more highly correlated with the true yield than were the model estimates, but the correlation for the larger counties were comparable, as shown by the weighted R^2.

Table 4. Comparison of Corn Estimates

Parameter	Model Estimates	Ratio Estimates
R^2	0.929	0.982
Weighted R^2	0.988	0.998
Sum(RMSE)	570	958
Estimates Produced (%)	98.4	92.7
Better Outliers (%)	94.5	0.0
Sum(Model RMSE Ests)	567	--------

Every other parameter of interest indicated that the model estimates were "better" than the ratio estimates. The significantly smaller Sum(RMSE) is particularly encouraging since this indicates that the model estimates are closer to the truth than the ratio estimates. The fact that the outliers produced by the model estimates are always as good or better than those produced by the ratio estimates is also very encouraging.

T-tests were performed on the squared differences between the two estimators for corn. The hypothesis that the model estimates are "better" than the ratio estimates was made, resulting in the test statistic assuming the following form:

H_o: Difference = 0

vs.

H_a: Difference < 0

In this case, 69 of 73 county differences (95%) were significantly less than zero at the $\alpha = 0.05$ level.

It's reasonable to conclude that the OSU model produced "better" corn estimates than the ratio estimator in this case.

The comparison of the model-produced RMSE estimates to the sample RMSE estimates was very interesting for corn. The Sum(Model RMSE Ests) found in Table 4 is nearly identical to the Sum(RMSE). The correlation between the average RMSE's from the 100 samples and the average model-estimated county RMSE was 0.777. Both of these comparisons indicate that, given a large sample, the model-based estimates of variation are reasonable.

For a more complete listing of statistics produced by the model see Table 7 in the Appendix which features various statistics for every county. Table 8 features complete statistics for the hypothesis tests for corn.

Conclusions and Recommendations

Use of the OSU algorithm offers several possible benefits:

1) Based on the preliminary testing, the program does appear to produce consistent results. This is a key element in the repeatability of an estimator. The fact that the estimates produced by the algorithm are

more consistent than standard ratio estimates across samples, as witnessed by the smaller sum(RMSE), indicates that the estimates produced by the algorithm are more repeatable.

2) Preliminary testing indicates that the model does a reasonable job of providing an indication of the variance of the estimates produced. As expected, the accuracy of these estimates improves as the sample size increases.

3) The goal of producing reasonable yield estimates for counties with no sample data available is achieved. This is demonstrated by the fact that more corn estimates are produced using the algorithm than with the ratio estimator. However, this was not the case for barley. This was probably a result of the inability to achieve convergence with the given sample and not a fault of the logic involved in the algorithm.

Before the program is implemented nationally, there are several matters that should be addressed:

1) In order to determine if the program works for a variety of commodities and cropping practices, additional testing should be conducted. To this end, beginning in January 2000 testing of this program was expanded to six offices. These offices include Ohio and Michigan, which have been involved in the past, as well as Tennessee, Mississippi, Colorado and North Dakota. The states were selected primarily to provide a wide range of commodities for testing. Additionally, Mississippi and Tennessee produce major commodities for which reliable administrative data are available at the county level, which will allow for a better evaluation of the estimates produced by the program. Finally, Colorado and North Dakota feature a wide variety of cropping practices which may or may not impact the estimates produced by the program.

2) As mentioned in the "Potential Problem" section, an algorithm should be developed to identify problem records in the event that the macro fails to converge to a single answer.

3) A method of obtaining previous year estimates must be developed. The statistician yield estimates are included with the published estimates which are forwarded to HQ, and this would be the logical place to obtain these estimates.

4) At some point, the OSU macro must be incorporated into the current county estimates summary system, so that the production of the yield indications is transparent to the office statistician.

5) Documentation of the technical details of the algorithm must be made available. The only available documentation of the macro currently available is a technical report written by OSU in 1995. Several changes have been made to the macro since that time, so a final guide or report should be produced which covers these changes. The group at OSU is currently working on a paper concerning the project, and this paper will likely provide the needed documentation.

6) The current version of the macro may not be the best small area estimation method available. Alternative methods or improvements to the current method will be examined under a cooperative agreement with Dr. Dan Griffith at Syracuse University. A method for deriving small area estimates for the Census of Agriculture has been examined by Dr. Griffith (1999), and this methodology may be extended to this project.

The following recommendations are made:

1) Evaluate the performance of the macro in additional field offices for a large variety of commodities. Feedback from ADP and commodity statisticians should be gathered which may lead to further improvements to the program. This evaluation will be conducted after the program is tested in six states.

2) Develop the procedures needed to fully incorporate the algorithm with the county estimates summary system. This includes:

a) The development of an algorithm to identify problem records.

b) The derivation of a method to automatically obtain previous year yield estimates.

3) Obtain and evaluate the forthcoming OSU paper as a potential source of documentation. The relevant sections of the paper should then be incorporated into the current county estimates documentation.

4) Assist Dr. Griffith with his examination of the program. Evaluate any suggestions made by Dr. Griffith which involve the yield program, and work to implement desired improvements.

References

Griffith, D. 1999. "A Methodology for Small Area Estimation, with Special Reference to a One-number Agricultural Census and Confidentiality: Results for Selected Major Crops and States," *NASS Research Report* RD-99-04. Washington, DC: Research Division, National Agricultural Statistics Service, U.S. Department of Agriculture.

Iwig, W.C. 1993. "The National Agricultural Statistics Service County Estimates Program", in "Indirect Estimators in Federal Programs", Statistical Policy Working Paper 21, Report of the Federal Committee on Statistical Methodology, Subcommittee on Small Area Estimation, Washington, D.C., 7.1-7.15

Stasny, E., Goel, P., Cooley, C., and Bohn, L. 1995. "Modeling county-level crop yield with spatial correlations among neighboring counties," *Technical Report No. 570.* Columbus, Ohio: Department of Statistics, The Ohio State University.

Appendix - Complete Results

The following paragraph is a description of the columns in Table 5 and Table 7. "Actual Population Yield" is the mean county yield from the generated population, and this number is considered the truth throughout this document. "Mean Model Yield" is the average of the model yields for the simulation samples. "Model RMSE" is the root mean square error calculated for the model yields from the simulation samples when compared to the "Actual Population Yields". "Mean Ratio Yield" is the average of the ratio estimates for the simulation samples. "Ratio RMSE" is the root mean square error calculated for the ratio yields from the simulation samples when compared to the "Actual Population Yields". "Mean Model RMSE" is the average of the root mean square errors produced by the model.

Table 5. Barley Results

County	Actual Population Yield	Mean Model Yield	Minimum Model Yield	Maximum Model Yield	Model RMSE	Mean Ratio Yield	Minimum Ratio Yield	Maximum Ratio Yield	Ratio RMSE	Mean Model RMSE
1	36.60	41.82	27.47	72.74	9.25	40.10	20.89	64.71	13.59	7.65
3	38.78	43.49	34.04	58.38	6.37	41.72	31.34	52.18	7.27	8.47
5	56.59	48.19	17.98	69.08	11.97	40.70	24.55	63.97	21.96	8.08
7	23.77	32.29	9.56	67.42	11.41	24.74	6.50	69.11	10.13	6.07
11	49.48	44.54	29.75	56.10	6.49	50.67	47.79	55.23	3.26	8.49
13	62.87	56.11	34.44	84.12	12.01	63.02	34.49	83.54	15.90	9.59
15	54.11	51.17	38.10	89.57	7.58	55.39	38.20	74.09	9.54	7.92
17	56.77	52.40	36.50	75.36	9.87	58.42	39.28	75.10	12.82	8.08
23	59.63	55.88	40.64	83.96	8.58	65.71	44.35	79.28	14.44	8.75
25	74.52	68.80	56.07	86.15	6.86	73.00	63.60	76.00	2.65	8.34
27	47.36	51.45	36.18	77.54	6.59	49.56	38.07	66.81	8.78	8.52
29	45.56	45.85	37.94	55.42	3.55	47.25	40.37	53.57	5.10	8.52
31	53.05	49.84	37.80	64.15	5.33	53.78	38.39	63.52	5.02	6.06
33	36.76	33.91	17.43	57.50	7.76	41.25	18.19	57.43	10.78	7.77
35	47.57	48.54	31.86	76.84	9.57	51.71	37.45	77.21	12.94	7.66
37	41.77	44.99	27.80	88.35	10.49	51.36	29.08	69.18	19.43	8.46
41	45.36	46.27	28.88	60.86	5.26	45.02	23.80	62.99	7.21	5.46
43	71.54	53.22	37.24	81.81	20.14	59.91	31.47	82.58	19.37	8.31
45	44.00	41.35	29.95	55.87	4.04	44.15	38.06	51.50	4.82	8.42
47	49.28	45.34	33.43	57.51	5.68	53.48	39.24	59.05	7.67	8.26
49	38.90	45.70	12.46	66.12	12.33	36.12	16.57	57.24	14.15	8.19
51	37.82	40.75	23.81	54.58	6.40	38.74	24.39	52.93	7.85	7.17
59	49.89	44.54	30.27	75.50	9.39	51.69	34.67	75.04	14.00	7.69
61	41.49	40.67	29.35	56.04	5.28	40.46	32.33	54.24	6.37	10.07
63	70.64	68.38	60.91	81.30	4.97	70.19	53.31	82.58	5.51	3.97
65	62.55	48.17	29.25	65.66	15.61	59.81	50.65	65.65	6.19	8.20
67	63.93	58.84	44.89	66.18	6.61	62.40	49.83	67.65	5.26	8.20

Table 5. Barley Results (Continued)

County	Actual Population Yield	Mean Model Yield	Minimum Model Yield	Maximum Model Yield	Model RMSE	Mean Ratio Yield	Minimum Ratio Yield	Maximum Ratio Yield	Ratio RMSE	Mean Model RMSE
69	47.57	50.27	39.03	67.29	6.38	48.59	40.18	63.90	8.37	8.77
71	47.73	45.77	29.25	59.28	5.25	47.03	27.25	59.12	6.57	8.48
73	54.04	55.57	25.45	76.28	6.87	51.75	26.40	71.32	9.72	6.43
75	60.13	54.95	35.66	80.89	9.22	56.49	36.05	74.71	12.90	8.19
81	42.25	47.40	27.85	62.81	8.86	42.32	30.63	52.43	6.44	7.69
87	42.87	42.23	18.44	66.05	5.80	41.08	12.34	67.44	10.33	6.57
91	50.06	55.00	46.09	70.24	6.50	53.88	48.59	58.67	6.04	8.52
97	38.13	41.81	27.30	60.65	7.02	37.19	30.10	56.26	5.96	9.10
101	51.60	41.74	27.22	53.81	11.81	53.53	39.60	62.54	7.52	8.82
103	54.20	55.50	41.25	76.23	5.70	50.13	43.02	57.10	6.45	8.11
105	62.01	51.05	33.33	67.13	12.48	62.04	57.80	65.23	2.84	9.06
107	42.44	42.19	12.47	62.69	7.11	41.32	4.57	60.51	11.25	7.19
109	44.13	44.70	35.50	53.31	3.52	43.43	32.30	54.28	3.93	3.99
113	47.90	49.09	31.31	62.42	6.39	48.11	19.89	62.40	10.46	8.44
117	57.31	50.15	36.03	74.50	9.45	53.94	26.10	72.70	11.47	6.80
119	43.67	49.86	29.23	62.17	8.40	46.21	20.61	59.32	12.30	8.05
123	54.53	50.50	38.95	76.92	6.11	55.39	38.80	65.11	6.94	7.46
127	53.89	54.21	42.16	67.04	4.96	53.34	46.45	60.84	5.91	8.35
129	30.96	44.33	12.90	68.00	16.06	38.13	12.52	69.80	16.83	10.79
131	41.86	46.81	36.82	59.98	8.36	44.90	37.76	60.07	10.01	9.39
133	46.40	48.96	15.53	69.19	9.54	43.18	17.94	69.19	12.08	8.06
137	36.41	34.86	10.29	55.64	7.63	35.53	10.97	53.81	11.09	8.02
139	41.48	41.64	28.41	59.96	5.31	37.63	18.60	56.46	12.55	8.27
141	43.30	40.41	31.62	48.76	4.38	41.93	29.09	51.41	4.89	4.71
147	45.08	43.72	29.30	74.47	9.85	51.86	33.23	76.06	16.73	8.35
151	47.55	50.74	34.10	63.00	5.68	47.14	33.94	62.28	6.13	4.52
155	70.55	64.76	52.69	83.81	8.47	69.40	54.55	89.39	8.03	7.51
157	59.55	61.04	45.46	77.44	5.55	60.23	48.19	73.05	5.79	6.57
159	39.59	39.27	25.86	62.11	5.59	41.83	27.92	58.07	9.00	8.04
161	50.27	42.73	25.63	64.17	9.45	49.18	22.88	63.98	9.69	7.01

Table 6. Barley T-Tests

County	N	Mean Difference	Variance of Mean Difference	T-Value	P-Value
1	49	-112.44	36,031	-4.147	0.000
3	47	-13.45	1,733	-2.216	0.016
5	38	-416.42	133,195	-7.034	0.000
7	59	36.23	51,904	1.221	0.887
11	21	35.54	1,195	4.712	1.000
13	31	-130.96	56,754	-3.061	0.002
15	47	-30.83	16,323	-1.654	0.052
17	39	-70.24	25,319	-2.757	0.004
23	27	-149.46	34,607	-4.175	0.000
25	27	48.08	1,745	5.980	1.000
27	46	-44.01	12,124	-2.711	0.005
29	40	-17.41	765	-3.981	0.000
31	62	3.41	2,015	0.598	0.724
33	58	-61.84	20,708	-3.273	0.001
35	40	-95.77	44,667	-2.866	0.003
37	28	-321.32	79,656	-6.024	0.000
41	63	-27.27	4,308	-3.298	0.001
43	49	33.23	129,564	0.646	0.739
45	26	-7.34	877	-1.264	0.109
47	32	-32.63	1,257	-5.206	0.000
49	35	-58.32	49,745	-1.547	0.066
51	54	-25.51	3,239	-3.294	0.001
59	41	-138.48	59,822	-3.625	0.000
61	24	-19.83	899	-3.241	0.002
63	63	-2.33	2,464	-0.372	0.356
65	31	224.52	16,734	9.664	1.000
67	28	13.36	3,032	1.284	0.895
69	29	-32.14	3,548	-2.906	0.004
71	47	-24.90	4,290	-2.606	0.006
73	61	-45.15	5,462	-4.771	0.000
75	30	-104.59	22,328	-3.834	0.000
81	41	45.10	4,708	4.209	1.000
87	58	-92.60	38,504	-3.594	0.000
91	27	10.94	2,588	1.117	0.863
97	18	14.99	3,715	1.044	0.844
101	27	88.44	24,852	2.915	0.996
103	40	-16.58	3,593	-1.749	0.044
105	22	197.16	20,351	6.482	1.000

Table 6. Barley T-Tests (Continued)

County	N	Mean Difference	Variance of Mean Difference	T-Value	P-Value
107	49	-99.99	55,791	-2.963	0.002
109	63	-4.85	581	-1.597	0.058
113	52	-88.25	28,452	-3.773	0.000
117	56	-50.32	27,310	-2.279	0.013
119	55	-93.95	26,359	-4.292	0.000
123	51	-13.28	2,723	-1.818	0.038
127	28	-16.39	1,638	-2.142	0.021
129	63	-20.99	46,329	-0.774	0.221
131	22	-39.78	7,853	-2.106	0.024
133	52	-71.80	53,221	-2.244	0.015
137	38	-82.86	23,132	-3.358	0.001
139	38	-155.54	39,015	-4.854	0.000
141	63	-5.25	945	-1.356	0.090
147	31	-208.72	148,433	-3.016	0.003
151	63	-1.65	3,626	-0.218	0.414
155	47	8.83	9,957	0.606	0.726
157	54	-6.27	2,869	-0.860	0.197
159	43	-62.82	10,724	-3.978	0.000
161	56	-6.80	10,448	-0.498	0.310

Table 7. Corn Results

County	Actual Population Yield	Mean Model Yield	Minimum Model Yield	Maximum Model Yield	Model RMSE	Mean Ratio Yield	Minimum Ratio Yield	Maximum Ratio Yield	Ratio RMSE	Mean Model RMSE
1	65.50	60.19	46.25	82.84	8.85	61.17	22.64	114.00	21.21	11.61
5	110.95	107.78	98.71	114.30	4.54	109.11	94.15	127.47	7.29	4.10
7	83.41	77.58	61.93	95.44	9.24	80.97	36.52	116.55	16.45	10.84
9	96.50	87.68	69.79	103.42	11.42	97.27	31.44	146.50	19.41	10.33
11	97.45	97.36	85.02	111.05	5.40	96.27	75.45	119.49	10.17	7.24
15	122.01	117.51	108.12	129.82	6.45	122.30	99.58	140.11	9.57	5.05
17	115.41	115.01	106.95	123.76	3.48	114.53	98.65	134.93	6.86	4.43
19	80.56	70.18	55.14	86.84	11.61	71.02	33.49	102.83	19.63	10.73
21	112.95	116.13	104.57	128.50	5.42	115.42	92.83	134.92	9.36	5.18
23	116.47	115.84	109.09	126.31	3.39	115.16	99.88	131.10	6.97	4.16
25	118.83	117.55	108.03	127.60	3.74	118.60	104.23	130.91	5.65	3.82
27	91.68	92.47	82.21	101.37	3.87	92.21	71.66	113.61	8.19	4.70
29	90.79	77.35	59.69	102.31	15.83	86.29	42.71	146.00	20.32	12.42
31	63.95	81.17	64.14	98.15	18.57	63.73	19.96	111.11	26.74	11.31
35	80.73	81.50	62.38	95.84	6.59	82.09	42.47	123.42	19.13	8.96
37	100.62	101.50	94.76	109.65	3.45	99.44	85.06	117.90	6.76	3.90
39	105.28	72.67	42.01	91.73	33.97	120.37	52.14	145.92	31.29	13.29
41	68.02	73.15	53.62	92.55	8.92	60.89	13.79	99.76	20.77	12.58
43	83.73	83.46	67.65	93.11	5.42	81.79	39.20	116.12	19.16	13.34
45	126.39	124.45	118.43	134.08	3.88	127.46	110.62	146.02	7.25	4.23
47	74.42	70.23	49.81	87.76	7.99	72.08	24.43	138.58	18.82	11.32
49	97.24	94.76	82.91	105.45	5.41	97.45	76.53	116.45	8.68	6.28
51	65.83	67.55	53.90	78.69	5.56	64.93	35.93	85.47	9.03	6.94
55	77.87	74.36	61.06	89.80	6.75	76.88	43.56	118.86	15.31	9.29
57	84.78	87.07	78.76	98.04	4.34	84.90	68.35	106.91	7.59	3.96
59	129.62	126.87	117.71	138.07	5.00	129.43	114.43	149.27	7.99	4.00
61	61.61	72.07	62.94	87.65	12.60	52.64	36.78	72.85	15.04	19.58
63	124.35	126.59	120.28	135.58	3.67	124.68	112.35	134.03	5.36	3.46
65	127.35	123.93	115.87	133.16	5.03	128.27	111.87	145.43	8.03	5.07
67	107.22	108.75	101.44	116.27	3.68	107.94	93.47	123.65	6.71	3.93
69	80.73	85.72	70.73	101.30	8.60	83.55	25.00	147.05	23.29	9.52
73	79.82	84.05	75.52	91.29	5.52	80.26	63.71	98.62	6.39	4.41
75	111.26	113.97	105.81	123.04	4.68	112.48	93.10	126.80	6.53	4.33
77	119.50	119.01	105.84	128.17	4.36	119.12	90.12	142.00	9.29	5.26
79	62.06	82.48	66.89	103.84	21.44	63.56	13.03	107.95	14.11	11.58
81	92.98	91.98	78.08	100.91	4.04	92.69	73.11	108.16	7.17	4.67
85	74.88	70.57	49.94	90.51	7.61	80.15	18.48	128.55	23.30	10.29
87	110.38	109.93	101.08	119.45	4.06	109.92	89.19	134.31	9.57	5.05
89	100.76	89.58	74.01	111.01	13.39	98.27	57.55	150.93	23.36	10.47
91	144.93	145.25	136.12	155.20	3.66	144.88	126.80	160.61	6.73	3.53
93	102.69	104.37	88.83	117.08	5.38	102.37	58.59	129.74	12.01	6.59

Table 7. Corn Results (Continued)

County	Actual Population Yield	Mean Model Yield	Minimum Model Yield	Maximum Model Yield	Model RMSE	Mean Ratio Yield	Minimum Ratio Yield	Maximum Ratio Yield	Ratio RMSE	Mean Model RMSE
95	82.56	95.23	61.16	112.15	19.71	93.94	62.26	126.03	27.18	24.03
99	116.96	111.33	97.21	123.97	8.33	114.64	73.54	139.04	13.41	7.20
101	105.74	90.21	73.74	113.11	17.28	105.27	72.14	129.45	12.65	9.47
105	89.15	89.74	72.37	101.76	4.93	88.08	66.55	106.50	8.02	7.56
107	95.15	88.68	76.72	104.87	8.14	94.83	74.22	131.25	10.40	6.08
109	96.07	90.57	69.01	111.13	9.51	93.57	59.50	136.96	16.24	10.80
111	91.99	95.15	81.14	105.89	6.06	94.17	68.93	127.11	11.17	6.23
113	100.65	97.64	83.85	117.37	7.03	103.08	72.13	133.00	12.15	8.76
115	152.63	150.40	139.47	158.68	4.77	152.61	136.43	169.24	6.76	4.07
117	98.52	94.31	86.00	101.73	5.55	97.92	82.15	118.33	7.37	4.27
119	68.01	84.01	65.46	109.37	18.35	71.90	24.55	126.21	25.28	14.41
121	108.33	102.24	90.25	117.69	8.35	106.44	78.16	145.33	12.81	7.04
123	84.75	88.70	78.42	97.68	5.59	84.20	61.26	107.77	7.82	5.17
125	79.70	88.20	76.01	101.28	10.14	81.69	40.49	118.68	16.69	9.24
127	83.82	82.50	71.65	96.04	4.91	83.22	65.12	116.52	10.24	6.09
129	84.14	87.32	72.77	101.50	6.74	85.03	47.79	120.61	12.32	8.64
133	81.02	85.98	73.20	100.76	7.66	81.44	46.37	112.08	12.46	8.20
135	64.14	72.06	56.31	85.83	10.36	71.26	14.20	112.14	22.89	13.44
137	81.01	73.71	57.76	106.73	10.20	76.96	20.57	121.41	20.70	11.66
139	108.37	105.95	97.86	114.69	4.49	106.95	83.44	134.18	9.44	4.32
141	78.24	80.04	65.34	93.89	5.97	81.04	38.51	138.30	16.44	8.96
145	100.31	102.66	95.96	110.61	3.76	101.02	85.95	116.16	6.02	3.46
147	102.43	100.64	88.94	113.34	5.10	101.46	73.19	129.09	8.63	5.36
149	98.52	102.44	92.50	113.51	5.73	99.67	68.35	118.29	9.56	4.33
151	128.91	130.18	118.55	136.29	3.39	129.52	116.06	140.00	4.69	3.40
153	66.07	77.66	55.42	87.82	14.49	78.17	35.11	103.66	27.44	14.47
155	86.68	86.88	78.74	94.70	3.47	84.90	68.06	106.72	8.01	4.19
157	125.01	126.32	116.88	134.38	3.70	125.44	96.05	139.87	7.16	3.63
159	99.80	99.08	82.98	111.59	4.60	101.73	77.09	122.80	9.35	6.04
161	119.94	118.40	109.91	127.48	3.84	119.40	105.38	135.32	6.36	4.20
163	105.28	111.18	91.91	131.90	8.67	106.81	53.48	180.28	20.59	9.77
165	92.96	81.88	68.51	99.75	12.66	80.66	38.36	129.34	21.01	10.57

Table 8. Corn T-Tests

County	N	Mean Difference	Variance of Mean Difference	T-Value	P-Value
1	85	-371.72	199,035	-7.682	0.000
5	100	-32.49	3,941	-5.176	0.000
7	99	-185.19	117,520	-5.375	0.000
9	94	-246.20	376,630	-3.890	0.000
11	100	-74.15	12,429	-6.651	0.000
15	100	-49.83	9,891	-5.010	0.000
17	100	-34.89	4,055	-5.479	0.000
19	96	-250.49	150,467	-6.327	0.000
21	100	-58.19	10,045	-5.805	0.000
23	100	-37.06	3,093	-6.663	0.000
25	100	-17.97	1,066	-5.506	0.000
27	100	-52.16	8,051	-5.813	0.000
29	95	-162.28	302,783	-2.874	0.003
31	82	-370.06	400,992	-5.292	0.000
35	100	-322.46	163,279	-7.980	0.000
37	100	-33.80	2,863	-6.317	0.000
39	27	175.25	741,216	1.058	0.850
41	94	-351.86	505,513	-4.798	0.000
43	51	-337.64	323,116	-4.242	0.000
45	100	-37.51	4,686	-5.479	0.000
47	95	-290.46	301,484	-5.156	0.000
49	100	-46.01	8,363	-5.031	0.000
51	100	-50.57	15,359	-4.080	0.000
55	100	-188.85	141,423	-5.022	0.000
57	100	-38.66	6,091	-4.953	0.000
59	100	-38.94	6,159	-4.962	0.000
61	29	-67.35	133,870	-0.991	0.165
63	100	-15.25	1,090	-4.620	0.000
65	100	-39.26	7,276	-4.602	0.000
67	100	-31.46	1,970	-7.087	0.000
69	100	-468.34	614,997	-5.972	0.000
73	100	-10.38	3,706	-1.704	0.046
75	100	-20.74	3,504	-3.503	0.000
77	100	-67.36	17,929	-5.031	0.000
79	90	260.55	193,731	5.616	1.000
81	100	-35.16	4,666	-5.147	0.000
85	80	-485.01	654,645	-5.362	0.000
87	100	-75.13	14,183	-6.309	0.000

Table 8. Corn T-Tests (Continued)

County	N	Mean Difference	Variance of Mean Difference	T-Value	P-Value
89	99	-366.54	263,013	-7.111	0.000
91	100	-31.82	3,034	-5.777	0.000
93	100	-115.39	53,737	-4.978	0.000
95	24	-350.24	296,052	-3.153	0.002
99	100	-110.39	74,947	-4.032	0.000
101	99	138.56	63,889	5.454	1.000
105	100	-39.96	7,866	-4.506	0.000
107	100	-41.85	31,872	-2.344	0.011
109	100	-173.51	99,502	-5.500	0.000
111	100	-88.09	28,398	-5.227	0.000
113	100	-98.23	33,099	-5.400	0.000
115	100	-22.96	3,261	-4.020	0.000
117	100	-23.50	6,083	-3.013	0.002
119	77	-302.29	531,309	-3.639	0.000
121	100	-94.28	51,034	-4.173	0.000
123	100	-29.97	11,018	-2.855	0.003
125	100	-175.63	106,766	-5.375	0.000
127	100	-80.82	19,490	-5.789	0.000
129	100	-106.45	47,081	-4.906	0.000
133	100	-96.46	50,629	-4.287	0.000
135	52	-416.43	445,936	-4.497	0.000
137	81	-324.34	468,937	-4.263	0.000
139	100	-69.07	21,735	-4.685	0.000
141	100	-234.49	225,085	-4.943	0.000
145	100	-22.10	2,273	-4.635	0.000
147	100	-48.48	13,310	-4.202	0.000
149	100	-58.53	18,981	-4.248	0.000
151	100	-10.55	521	-4.624	0.000
153	24	-543.09	132,995	-7.296	0.000
155	100	-52.08	6,531	-6.445	0.000
157	100	-37.50	10,418	-3.674	0.000
159	100	-66.29	13,592	-5.686	0.000
161	100	-25.65	2,164	-5.515	0.000
163	99	-348.68	573,585	-4.581	0.000
165	95	-281.26	350,690	-4.629	0.000